

University of Groningen

## Generating personalized advice for schizophrenia patients

Emerencia, A.; van der Krieke, Lian; Sytema, S.; Petkov, N.; Aiello, M.

*Published in:*  
Artificial Intelligence in Medicine

*DOI:*  
[10.1016/j.artmed.2013.01.002](https://doi.org/10.1016/j.artmed.2013.01.002)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2013

[Link to publication in University of Groningen/UMCG research database](#)

### *Citation for published version (APA):*

Emerencia, A., van der Krieke, L., Sytema, S., Petkov, N., & Aiello, M. (2013). Generating personalized advice for schizophrenia patients. *Artificial Intelligence in Medicine*, 58(1), 23-36.  
<https://doi.org/10.1016/j.artmed.2013.01.002>

### **Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### **Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*



## Generating personalized advice for schizophrenia patients

Ando Emerencia<sup>a,\*</sup>, Lian van der Krieke<sup>b</sup>, Sjoerd Sytema<sup>b</sup>, Nicolai Petkov<sup>a</sup>, Marco Aiello<sup>a</sup>

<sup>a</sup> University of Groningen, Nijenborgh 9, 9747 AG Groningen, The Netherlands

<sup>b</sup> University Medical Center Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands

### ARTICLE INFO

#### Article history:

Received 23 December 2011

Received in revised form

11 December 2012

Accepted 8 January 2013

#### Keywords:

Deductive reasoning

Ontology

Knowledge-based web information system

Routine patient assessments

Schizophrenia

### ABSTRACT

The results of routine patient assessments in psychiatric healthcare in the Northern Netherlands are primarily used to support clinicians. We developed Wegweis, a web-based advice platform, to make this data accessible and understandable for patients.

**Objective:** We show that a fully automated explanation and interpretation of assessment results for schizophrenia patients, which prioritizes the information in the same way that a clinician would, is possible and is considered helpful and relevant by patients. The goal is not to replace the clinician but rather to function as a second perspective and to enable patient empowerment through knowledge.

**Methods:** We have developed and implemented an ontology-based approach for selecting and ranking information for schizophrenia patients based on their routine assessment results. Our approach ranks information by severity of associated schizophrenia-related problems and uses an ontology to decouple problems from advice, which adds robustness to the system, because advice can be inferred for problems that have no exact match.

**Results:** We created a problem ontology, validated by a group of experts, to combine and interpret the results of multiple schizophrenia-specific questionnaires. We designed and implemented a novel ontology-based algorithm for ranking and selecting advice, based on questionnaire answers. We designed, implemented, and illustrated Wegweis, a proof of concept for our algorithm, and, to the best of our knowledge, the first fully automated interpretation of assessment results for patients suffering from schizophrenia. We evaluated the system vis-à-vis the opinions of clinicians and patients in two experiments. For the task of identifying important problems based on MANSA questionnaires (the MANSA is a satisfaction questionnaire commonly used in schizophrenia assessments), our system corresponds to the opinion of clinicians 94% of the time for the first three problems and 72% of the time, overall. Patients find two out of the first three advice topics selected by the system to be relevant and roughly half of the advice topics overall.

**Conclusions:** Our findings suggest that an approach that uses problem severities to identify important problems for a patient corresponds closely to the way a clinician thinks. Furthermore, after applying a severity threshold, the majority of advice units selected by the system are considered relevant by the patients. Our findings pave the way for the development of systems that facilitate patient-centered care for chronic illnesses by automating the sharing of assessment results between patient and clinician.

© 2013 Elsevier B.V. All rights reserved.

### 1. Introduction

Schizophrenia is a mental disorder that affects approximately 1% of the population. The illness is characterized by psychoses, which are episodes involving a loss of contact with reality. The symptoms of the illness are caused by impaired processing of

information in the brain in combination with gene–environment interactions [1].

Current schizophrenia treatment in the Northern Netherlands is centered around Routine Outcome Monitoring (ROM) to assess schizophrenia patients. In recent years, ROM has become increasingly important as part of a growing belief in the need for standardization in order to evaluate and improve patient care. A ROM assessment for a patient is conducted every 6 months or every year. These assessments involve physical fitness tests, as well as a number of questionnaires that assess psychiatric and psychosocial problems, satisfaction, and care needs. The ROM protocol makes use of several questionnaires such as the Health of

\* Corresponding author. Tel.: +31 50 363 3984; fax: +31 50 363 3800.

E-mail addresses: [a.c.emerencia@rug.nl](mailto:a.c.emerencia@rug.nl) (A. Emerencia),

[j.a.j.van.der.krieke@umcg.nl](mailto:j.a.j.van.der.krieke@umcg.nl) (L. van der Krieke), [s.sytema@umcg.nl](mailto:s.sytema@umcg.nl) (S. Sytema),

[n.petkov@rug.nl](mailto:n.petkov@rug.nl) (N. Petkov), [m.aiello@rug.nl](mailto:m.aiello@rug.nl) (M. Aiello).

the Nation Outcome Scales (HoNOS) [2] and the Manchester Short Assessment of Quality of Life (MANSA) [3].

The results of a ROM assessment form the basis for a long-term treatment plan that is determined in a meeting between patient and clinician. These meetings take place roughly six weeks after an assessment. During the meeting, a treatment plan is formulated that is followed until the next assessment.

There is increasing concern that patients are not sufficiently engaged in these meetings, because they are not always adequately prepared to have a discussion. Patients have no direct access to the assessment results prior to the meeting and hear these results only through their clinician. This scenario creates an inequality, wherein the patient is highly dependent on the expertise of the clinician and cannot participate fully in medical decision making. In recent years, the ethics of such medical paternalism have been called into question [4].

To better prepare patients for meetings with their clinician, tools have recently been developed to support *shared decision making* [5,6], which is considered an ethical imperative [7]. Shared decision making is an approach in which patient and clinician are equal participants in deciding the treatment plan. Moreover, the approach emphasizes that patients should have access to the same information regarding their (mental) health as the clinician [8]. Shared decision making is widely in use and has proved clinically successful for most chronic illnesses.

So far, however, sharing healthcare information with the patient in a direct and unsupervised manner, as part of shared decision making, has not been applied in terms of schizophrenia patients. Moreover, to the best of our knowledge, there has been no research on the automated conversion of assessment results into relevant information for schizophrenia patients. There are two main reasons for this. First, clinicians have traditionally subscribed to the belief that they need to protect their patients against potentially disturbing outcomes. Second, tools that facilitate shared decision making for schizophrenia patients require careful development because schizophrenia patients have special needs regarding the presentation of information, for example, via a simply structured and calm website using text for a low reading age [9], that is, using text without difficult words.

The main contribution of this paper is to show that a fully automated explanation and interpretation of ROM assessment results for schizophrenia patients that prioritizes the information in the same way that a clinician would is possible and is considered helpful and relevant by patients. This work forms an important step towards implementing shared decision making as part of the standardized approach in schizophrenia treatment.

In this paper we will present, evaluate, and explain our web application called Wegweis, which can perform an automated explanation and interpretation of ROM assessment results. Wegweis was designed in iterations using feedback from patients and in cooperation with clinicians from all four mental health institutions in the Northern Netherlands (GGZ Drenthe, GGZ Friesland, Lentis, and UCP). Wegweis supports shared decision making by providing patients with their assessment results and an interpretation thereof in the form of personalized advice.

Since not every patient is eager to be confronted with the problems of their illness, Wegweis offers solution-oriented information. In order to make the website attractive for patients, the information is presented in the form of advice, personalized suggestions, helpful tips, and information. The advice consists of information derived from evidence-based research (e.g., the Dutch Multidisciplinary Guideline for Schizophrenia), clinical expertise, and patient experiences. For example, the contents of the advice units range from recommending nearby fitness centers and patient organizations, to providing information about medication side effects and locally available cognitive behavioral therapy modules.

To the best of our knowledge, Wegweis is the first web application that is able to rank information as experienced clinicians do and in a way that is considered helpful by schizophrenia patients, as we will show in this paper. In it we will explain how we designed and implemented an ontology-based approach to reasoning over background knowledge and to determining the applicability and specificity of relevant information for a patient. Ranking information simplifies navigation for a patient, since the most relevant information is likely to be on the first few pages of the results.

With the availability of Wegweis as a web application, patients can access its information at any time, and without pressure or supervision. Patients should be given access to Wegweis prior to meeting with their clinician. Wegweis encourages patients to bring their own point of view to the discussion, thereby making patient and clinician equal participants in deciding the treatment plan.

The rest of the paper is organized in the following way: Section 2 gives an overview of related work; Section 3 explains the system design of Wegweis; Section 4 explains the user interface; Section 5 details the problem ontology; and Section 6 presents the algorithm for selecting and ranking advice for a patient. We evaluate the system in two experiments reported in Section 7, discuss the results in Section 8, and present some conclusions in Section 9.

## 2. Related work

There are numerous examples of ontology-based applications in healthcare. For example, ontologies are used in the middleware of pervasive health systems for monitoring patients and managing alerts [10] and for generating clinical reminders for clinicians [11]. Another example is TrialX, a web application that uses its own ontology to interpret and evaluate data stored in personal health records in order to match patients to clinical trials [12]. More closely related, SEMPÉR is an interactive web-based platform that assists patients to self-manage work-related disorders and alcoholism, and uses ontologies for query expansion in text mining in documents [13]. Kuriyama and colleagues [14] developed an application for mobile devices for collecting and sending lifestyle data that are used to display health advice in a web application. They use an ontology to suggest exercises based on the goals of the patient.

In relation to other ontology-based applications in healthcare, our application (Wegweis) is novel because it is the first application that shows information originally intended for clinicians (assessment results) to schizophrenia patients, and uses an ontology to automate the translation from results to information. This automated translation is an important step in implementing one of the core requirements of shared decision making, the sharing of medical information, at low operational costs.

While patient-supporting web applications are already in use for mental illnesses such as anxiety, depression, and addiction [15], for schizophrenia and other severe mental illnesses, less has been achieved thus far [16–18].

In Finland, Välimäki and colleagues [18] have developed the Mieli.Net portal, a patient-centered computer-based support system for patients with schizophrenia spectrum psychoses. It aims to support self-management by offering (i) information on treatment, support, and rights; (ii) a channel for peer support; (iii) a tool for counseling; and (iv) interaction with clinicians by means of a question-and-answer column. A prototype was developed and has been evaluated by patients and healthcare staff. Both nurses and patients were able to work with the system [18–20]. Patients were able to access services and find relevant information [19], and they report their satisfaction with the system [21].

In the Netherlands, two recent initiatives have been launched aimed at enabling empowerment of schizophrenia patients. The first is “Eigen regie bij schizofrenie” (translation: personal control

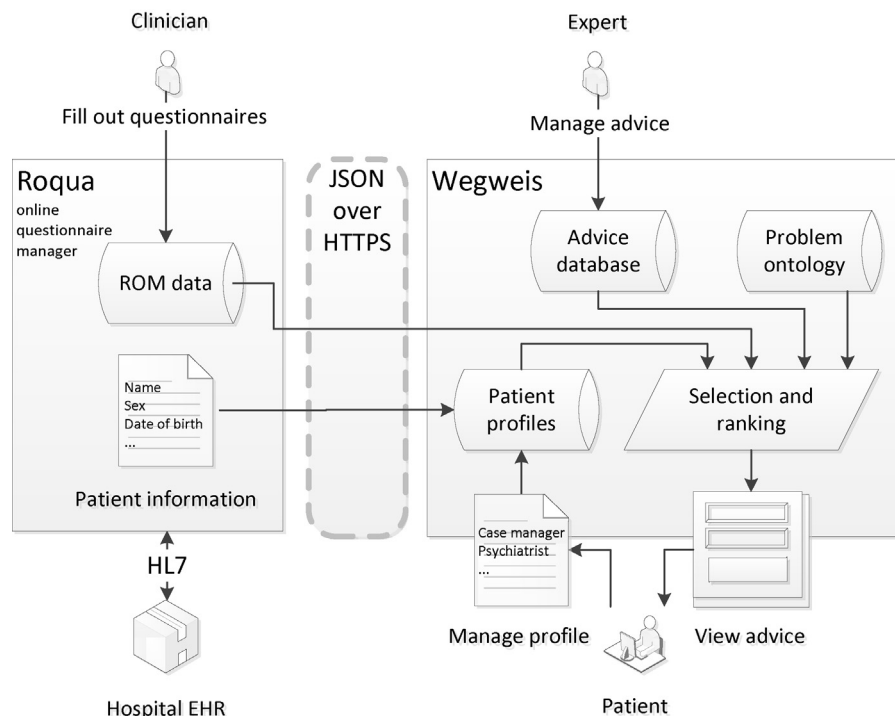


Fig. 1. Flow of information for selecting and ranking advice in Wegweis.

over schizophrenia), a website to support patients in their self-management [22]. It offers tools for scheduling appointments, checking medication, viewing the treatment plan, sharing experiences, and requesting services. Clinicians can use the website to monitor the condition of patients and detect problems early. The second is SamenKeuzesMaken.nl (translation: making decisions together), a website that is modeled after a program of Deegan and colleagues [23] that implements the concept of shared decision making [24]. It offers information about recovery, videos portraying experienced patients, a questionnaire in preparation for meeting the clinician, and links to informational websites. We note that there is no true sharing of information here, since the patient fills out a separate questionnaire on the website and does not gain access to the assessment results that his/her clinician has.

While there are other web applications for schizophrenia patients that support shared decision making, they do not support the direct sharing of assessment information. With Wegweis, a direct translation becomes possible through applying ontological reasoning, as we will explain in this paper. Wegweis can rank and personalize information for individual patients. This functionality can also be abstracted and applied to existing self-management websites in order to make them more personalized and easier to use for patients.

### 3. Wegweis system design

To facilitate its main functionality of generating and showing advice to patients, Wegweis retrieves information from external services and has an interface for experts to manage the advice.

Retrieving information from external services is illustrated in Fig. 1. This figure shows how Wegweis retrieves patient information and ROM data from Roqua, an online questionnaire manager used by mental health institutions in the Northern Netherlands [25]. Roqua is used by clinicians and interfaces with electronic health records at mental health institutions. Thus, Wegweis interfaces only indirectly with the electronic health records. Roqua interfaces with the EHRs using HL7, a communications

standard used in healthcare applications [26]. The communication between Roqua and Wegweis uses JSON [27] over HTTPS.

Fig. 1 also shows that patients can view their advice, and that experts can manage the advice units. Patients view advice based on an advice selection and ranking process that uses questionnaire answers, patient information, and a problem ontology. We note that all domain knowledge is isolated in the problem ontology, so the approach used by Wegweis is not necessarily schizophrenia-specific. Wegweis has an interface for experts to manage the advice units. The advice units that we used for our experiments (Section 7) were written with an emphasis on keeping the text simple and to the point, and were validated by psychiatrists, psychologists, and patients. The user interface for managing advice units is described in the next section.

Before patients can view their advice, they need to have an account with Wegweis. We created a plug-in for Roqua that allows clinicians to send patients an invitation for Wegweis. Sending an invitation also sends a request to Wegweis to create an account for the patient, and allows Wegweis to retrieve ROM data and patient information for that patient through Roqua. After the invitation is sent, the patient decides whether or not to respond to the invitation. The invitation e-mail links to an account-creation page in Wegweis that is authorized to create an account linked to the information of that particular patient. On the account-creation page, the patient can optionally provide Wegweis with the names of his/her psychiatrist and case manager, which are used to personalize the advice texts. Once the account has been created, the patient is instructed to click on “My Advice” which will immediately show the advice that our system has selected, based on the assessment results. In this paper we explain how our system selects and ranks advice for patients.

### 4. Wegweis user interface

Schizophrenia patients have specific needs regarding the content, structure, and layout of a website [9]. They frequently have cognitive problems, such as concentration problems, as a result

Fig. 2. Part of the “My Advice” page in Wegweis.

of the illness and side effects of medication. Rotondi and colleagues [28] showed that for people with severe mental illnesses, best practices are to keep the navigation simple, to keep words and phrases simple, to avoid having too much text on one page, and to refrain from using flashing or otherwise distracting elements.

We designed and implemented a way to display advice that respects these limitations. Fig. 2 shows part of the “My Advice” page, listing the first page of advice for a patient. This page originally contained Dutch text; shown here is a translation. The advice on the page is divided into three sections. We call these sections *advice units*. Each advice unit has a title, in bold, that represents the problem area (e.g., “Is school or work not going so well?”) and two or three solutions, shown in the gray boxes. Note that these solutions are just single lines of text. By clicking these lines, interested readers can open up more information. These expanded contents can again contain collapsed elements. Thus, we gradually show more information to the patient by revealing small chunks of text at a time. This interface was found to be usable by most schizophrenia patients in our usability study [29].

Wegweis employs aspects of *personalization* to appeal to patients. Personalization in web applications can be defined as any action that tailors the web experience to a particular user or set of users [30]. Wegweis implements two levels of personalization in the process of generating advice for patients. First, the selection of advice units and the order in which they are presented depends on the ROM data of a patient, and is therefore personalized. This process of selecting and ranking advice units is part of the main contribution of this paper, and is explained and evaluated in Sections 6 and 7. Second, the contents of the advice units can be made to appear more personal by including certain variables. These variables are evaluated at run-time in the context of the patient. For example, when we use the variable `case_manager` or `psychiatrist` in the advice contents, the patients will see the actual name of their practitioner instead. This second level of personalization is implemented by simply locating all occurrences of variables and

**Advice problems:** click a problem to toggle selection. All selected problems are always visible (expanded). Questionnaire bindings help.

Fig. 3. The expert interface for adding an advice unit.

replacing them with the corresponding information from patient profiles.

While the interface for managing advice units in Wegweis (shown in Fig. 3) is based on an existing CMS framework called BrowserCMS [31], we implemented additional functionality to facilitate writing advice units. Fig. 3 shows how the problems that are associated with an advice unit (i.e., the problems that can trigger an advice unit) are selected from a tree view. The advice contents are written in the Liquid templating language [32]. We chose a lightweight templating language, since it allows people without a technical background to easily create HTML content. We extended the Liquid syntax to allow for customized variables (`case_manager` and `psychiatrist`) and scopes (collapsed text, tips, warnings, quotes, and notes). The advice units can embed audio clips, video fragments, as well as other advice units (e.g., when reusing common texts). We also added a live preview with syntax checking for the advice contents, to avoid common errors. Advice units can be added on-the-fly and changes propagated immediately. The advice pages load without noticeable delay, because intermediate stages of the advice unit selection process are cached and embedded content is loaded asynchronously. The implementational details of the staged caching process fall outside the scope of this paper.

## 5. Problem ontology

The advice ranking and selection process in Wegweis is based on questionnaire items (i.e., the questions of a questionnaire), which are handled individually. This individual treatment contrasts with the common interpretation of schizophrenia questionnaires. Commonly, schizophrenia questionnaires are interpreted through mean or summation scores of multiple items [2,3]. We chose to handle



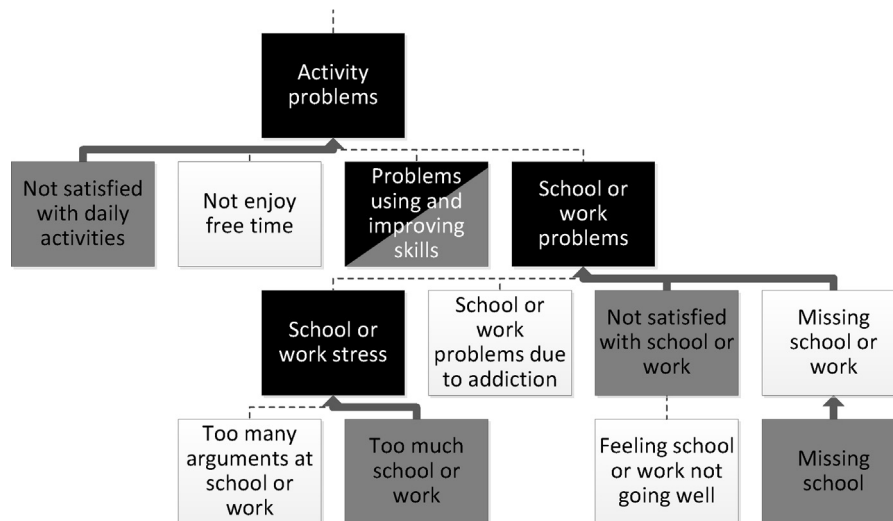


Fig. 4. Part of the ontology.

each item individually to keep information loss at a minimum, on the assumption that each item identifies a distinct problem. Hence, we use the terms “questionnaire item” and “problem” interchangeably.

Our approach for the individual treatment of questionnaire items involves (i) identifying a schizophrenia-related problem for each item and (ii) interpreting the answer as a measurement of the severity of that problem for a patient. This two-step process transforms a filled-out questionnaire into a list of problems and severities. The second step in this process (i.e., interpreting a questionnaire answer as a problem severity) is detailed in the next section, where we show how the list of problems and severities selects and ranks the advice units for patients. The first step (i.e., associating questionnaire items with schizophrenia-related problems) and the problem ontology used therein are explained in the remainder of this section.

Recognizing questionnaire items as individual problems creates 97 problem variables for the four questionnaires that we consider (16 for MANSA [3], 12 for HoNOS [2], 24 for CANSAS-P [33], and 45 for OQ-45 [34]), some of which we found to be very similar. For example, item 11 of the OQ-45 questionnaire is associated with the problem called *AlcoholAbuse*, while item 3 of the HoNOS questionnaire is associated with the problem called *AlcoholOrDrugAbuse*. Since these two problems are semantically similar, it is likely that an advice unit that applies to one of them also applies to the other. Associating an advice unit with problems would be tedious if we had to determine applicability for all problems of all questionnaires manually.

In order to take advantage of the similarities that exist among the problems identified, we created a *problem ontology*, which imposes a hierarchy on the problems and allows us to identify groups of problems with similar semantics. In contrast to the traditional approach of interpreting schizophrenia-related questionnaires (which considers the summation of the severities of a group of related questionnaire items), our approach considers the maximum severity. Thus, in our approach, any individual problem that is severe enough can trigger advice. Hence, we can tailor the advice for a patient, based on individual problems.

The problem ontology decouples the questionnaire items from the advice units and thereby simplifies the process of associating an advice unit with problems. The decoupling is due to the fact that we associate questionnaire items and advice units with problem concepts rather than with each other. The simplification in advice unit association is due to the knowledge stored in the ontology that

allows us to associate an advice unit with those problems that represent groups of semantically similar problems, rather than having to determine all applicable problems manually.

In our ontology, the schizophrenia-related problems are the only *concepts* and their hierarchy is the only *relationship*. This relationship, called the *is a* relationship, is a partial order (i.e., relations are reflexive, antisymmetric, and transitive) that denotes specificity. Essentially, the inferred relationships form a tree with root node *Problems* that branches out into increasingly specific problems. Thus, every child node is a more specific problem concept of its parent node. For example, in our ontology, the node *Fatigue* has the following ancestors (listed in reverse hierarchical order): *NegativeSymptoms*, *PsychoticProblems*, *PsychicProblems*, and *Problems*. From the properties of our ontology, we deduce that the applicable advice for an active problem concept (i.e., a problem affecting the patient) consists of the advice associated with the problem concept or with any of its ancestors.

In our approach, the ontology is traversed in reverse hierarchical order to find advice in cases where an active problem concept is not associated with any advice units. This process is illustrated in Fig. 4. This figure shows part of the ontology as a tree with problem concepts as nodes and the *is a* relationship as edges. Furthermore, in this figure, nodes with a black background are associated with advice units, nodes with a gray background are active nodes (i.e., associated with a questionnaire item that was answered above a certain threshold), and nodes with a white background are inactive and can be ignored. We make no distinction between leaf nodes and other nodes, i.e., any node can be associated with advice units, with questionnaire items, or with both. The arrows in Fig. 4 indicate the paths from active nodes to their first ancestor that is associated with advice and show how advice for certain questionnaire problems is found higher up in the ontology. For example, advice that is associated with the *School or work problems* node will be triggered with the maximum problem severity of the questionnaire items associated with the *Not satisfied with school or work* and *Missing school* nodes. We cover the algorithm for selecting and ranking advice units in more detail in the next section.

We opted to create a new ontology rather than using an existing ontology, because we found that existing ontologies did not cover some of the problem concepts that we identified. Our idea was that the problem ontology should represent the full spectrum of problems that can affect a schizophrenia patient. The recommended approach for using ontologies in healthcare applications is to use an existing medical ontology such as SNOMED-CT [35]. However, we

found that existing medical ontologies have no equivalent for some of the identified problem concepts. This is because some of the identified problem concepts are not medical in nature or not associated with the patient. For example, item 2 of the MANSA questionnaire asks whether the patient is satisfied with his/her residence, which in our ontology is associated with the `NotSatisfiedWithResidence` problem concept. This concept has no equivalent in existing medical ontologies, since the problem is not medical in nature and (arguably) not associated with the patient but with his/her residence.

The primary argument for using an existing ontology is to facilitate interoperability (i.e., exchanging data with other systems), which can still be achieved with our approach. In our case, interoperability would refer to the importing and exporting of patient summaries. With our custom ontology, we can still achieve interoperability by associating (a subset of) the problem concepts with a standardized ontology, such as SNOMED-CT, in an ontology mapping. With such an ontology mapping, we can use the same algorithms that we designed for finding the most relevant advice to find the most relevant concepts that exist in a standardized ontology, thus allowing for interoperability with other systems that use the same ontology.

We constructed the problem ontology for Wegweis with the help of a psychiatrist and a psychologist. These professionals identified relationships among problem concepts and indicated groups of problems, to which the same advice would apply. We incorporated their assessments into the structure of the problem ontology. This ontology (including the associations with advice units and questionnaire items) was validated by ROM experts and clinicians. They stated that they had studied the ontology and did not find any abnormalities. Furthermore, they noted that the reasoning applied in the hierarchy was sound and made intuitive sense.

We implemented the problem ontology using Protégé [36] in OWL, the Web Ontology Language [37]. Expressed in OWL terminology, the problem concepts are `Classes` and the relationships are defined using `SubClassOf` axioms. The inferred hierarchical structure of the ontology is the result of running the HermiT 1.2.2 Reasoner on the ontology in Protégé. The inferred ontology is exported to an OWL file that is parsed by Wegweis. In addition to the problem concepts and their hierarchy, the ontology also stores the associations between questionnaire items and problem concepts, but it does not store the associations between advice units and problem concepts. Our reasoning for this design is that both the problem concepts and the questionnaire items make sense to domain experts (i.e., they make sense outside the context of Wegweis), while advice units are objects specific to Wegweis. The associations between advice units and ontology concepts are stored in the database of Wegweis. Wegweis identifies ontology concepts by their name and continuously monitors the OWL files to avoid inconsistencies. For example, if a problem concept was removed from the problem ontology, then any advice unit associated with this problem concept should be updated to reflect that it can no longer be activated by said problem concept. In contrast, the associations between questionnaire items and ontology concepts are part of the ontology and are modeled in OWL as `AnnotationAssertion` axioms with questionnaire items represented as `Literals` (e.g., `Mansa_1`, `HoNOS_5`). Our ontology is available online [38].

## 6. Selecting and ranking advice

Since having too much text on one page can overwhelm the patient [9], Wegweis shows only three advice units per page. Therefore, the order in which these advice units are listed is important. We let the order of advice units be determined by the inferred severity of the problems associated with them. We use

no exclusion criteria for advice, since we consider leaving out key advice more harmful than giving too much advice. In our experiments, we assessed the validity of our approach (see Section 7). We first introduced the algorithms for implementing our approach in [39], without an evaluation. Everything about these algorithms, including the design, terminology, and implementation, was done by us.

### 6.1. An algorithmic overview

Fig. 5 gives an overview of our approach for transforming the answers of a patient for a certain questionnaire into a sorted list of advice units. The *problem severities* shown in the overview are the result of a preprocessing step in which the raw questionnaire answers are normalized. Thus, after the preprocessing step, we have the problem severities for the problem concepts that are associated with the questionnaire items of the filled-out questionnaire. For these problem concepts and for all their ancestors in the ontology, we calculate a similar metric that we call the *activation strength*, which combines problem severity with specificity, as we will explain in this section. Finally, we convert a list of problem concepts and their activation strengths into a list of advice units and their priorities. We define the *priority* of an advice unit as the maximum activation strength of the problems that are associated with the advice unit. The result is a list of applicable advice units and their priorities. These priorities are then used to sort the applicable advice, and this sorted list of advice units then forms the contents of the “My Advice” pages such as the one shown in Fig. 2. The remainder of this section describes the above steps in more detail, with the help of pseudocode and a sample run case.

In the preprocessing step of our approach, we convert questionnaire answers into problem severities. We define the term *problem severity* to denote the normalized questionnaire answer such that 0 and 1 denote the least and most severe answer option, respectively, and values for intermediate strata follow from linear interpolation at equidistant intervals. For example, most items of the MANSA questionnaire are rated on a seven-point satisfaction scale, from 1 = “Couldn’t be worse” to 7 = “Couldn’t be better”. Thus, the problem severity corresponding to answer 1 is 1, since it denotes the most severe condition, and analogously the problem severity corresponding to answer 7 is 0. Likewise, an item answered with 2 = “Displeased” translates to a problem severity of  $\approx 0.833$ . Translating questionnaire answers into problem severities in this way is possible because we found that the schizophrenia questionnaires that we considered had the same structure. In this structure, the questionnaire items relate to some problem or condition, and the answers are an indication of how much the problem affects the patient and are expressed on a rating scale with a certain number of strata. These linear rating scales allow for a straightforward normalization to unit range.

The core of our approach, shown in Fig. 5, is our *advice unit priority algorithm*, a two-step process that converts problem severities into advice unit priorities. As we explained earlier, the problem severities map problems (associated with questionnaire items) to severities (the normalized questionnaire answers). Our algorithm consists of two steps: (i) calculating the activation strengths and (ii) using the activation strengths to calculate the advice unit priorities. We will describe these steps next.

### 6.2. Calculating the activation strengths

In the first step of our *advice unit priority algorithm*, we convert problem severities into activation strengths. We define *activation strengths* as *(level, severity)* tuples that are ordered lexicographically by highest level first and by highest severity second. For

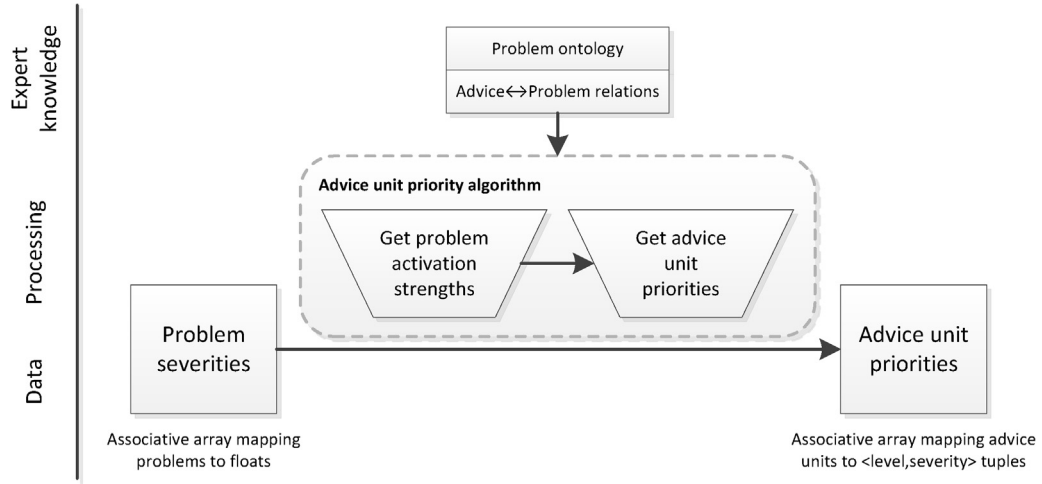


Fig. 5. An overview of our approach for using problem severities to rank advice units.

example, the following list of activation strengths appears sorted in order:  $\langle 0, 0.33 \rangle$ ,  $\langle -1, 0.83 \rangle$ ,  $\langle -1, 0.44 \rangle$ . The activation strength for a problem  $p$  is calculated as the maximum augmented activation strength of  $p$  and its descendants, where the augmentation for a descendant  $q$  of  $p$  consists of decreasing the specificity for every advice unit that applies to  $q$  but not to  $p$ . For example, imagine that we want to calculate the activation strength of the School or work problems node in Fig. 4, with the following nodes being active: Missing school with problem severity 0.25, Not satisfied with school or work with problem severity 0.50, and Too much school or work with problem severity 0.75. Now, the activation strengths of these nodes from the point of view of the School or work problems node are  $\langle 0, 0.25 \rangle$  for Missing school,  $\langle 0, 0.50 \rangle$  for Not satisfied with school or work, and  $\langle -1, 0.75 \rangle$  for Too much school or work. The Too much school or work node has a lower level, since there is an advice

unit (associated with the School or work stress node) that applies to the Too much school or work node but not to the School or work problems node. Thus, the activation strength of the School or work problems node is  $\langle 0, 0.50 \rangle$ , which is the maximum augmented activation strength of itself and its descendants, since the tuples are ordered lexicographically by highest level first and by highest severity second.

A description in pseudocode for this step is the GETPROBLEMACTIVATIONSTRENGTHS algorithm shown in Fig. 6. This algorithm starts by initializing  $P$  to be the set of all problem concepts in the ontology and  $T$  to be a mapping of problems to activation strengths, which are initialized as tuples of problem severities with level 0 for the nodes associated with active questionnaire items. In the algorithm,  $T$  and  $A$  hold intermediate results, while  $B$  is eventually returned. The outer loop traverses over all nodes in  $P$  by selecting the leaf nodes of  $P$  in every iteration and removing them from  $P$  afterwards.

**Algorithm 1:** GETPROBLEMACTIVATIONSTRENGTHS( $V$ )

**Input:** associative array  $V$  mapping problems to problem severities (floats).

**Data:** ontology functions `all_problems` and `descendants`.

**Output:** associative array mapping problems to  $\langle level, severity \rangle$  tuples, for all triggered problems.

```

 $P \leftarrow \text{all\_problems}()$ 
 $B \leftarrow$  empty associative array
 $T \leftarrow$  empty associative array
 $A \leftarrow$  empty associative array
for each problem  $p \in V.\text{keys}$ 
  do  $T[p] \leftarrow \langle 0, V[p] \rangle$ 
while  $P$  is not empty
   $N \leftarrow \text{GETLEAFNODES}(P)$ 
  for each problem  $p \in N$ 
    for each problem  $q \in \text{descendants}(p)$ 
      do if  $T[q]$ 
        then  $T[p] \leftarrow \max(T[p], T[q])$ 
      if  $T[p]$ 
        then  $B[p] \leftarrow T[p]$ 
        remove  $p$  from  $P$ 
   $T, A \leftarrow \text{UPDATEPROBLEMLEVELS}(N, T, A)$ 
return ( $B$ )

```

Fig. 6. The GetProblemActivationStrengths algorithm.



**Algorithm 2:** GETLEAFNODES( $P$ )

**Input:** set of problems  $P$ .  
**Data:** ontology function `descendants`.  
**Output:** the subset of problems that are relative leaf nodes.

$L \leftarrow$  empty set  
**for each** problem  $p \in P$   
  **do**  $\left\{ \begin{array}{l} \text{if } (\text{descendants}(p) \cap P) \text{ is empty} \\ \text{then add } p \text{ to } L \end{array} \right.$   
**return** ( $L$ )

Fig. 7. The GetLeafNodes algorithm.

In the inner loop,  $T[p]$  is set to the maximum  $T$  value of  $p$  and its descendants, and if this value is not null, then it is copied to  $B[p]$ . When all leaf nodes in an iteration have been considered,  $T$  and  $A$  are updated to account for advice given in the iteration.

The algorithm makes use of the GETLEAFNODES function, which is shown in Fig. 7. This function returns the subset of relative leaf nodes within a given set of nodes  $P$ . The relative leaf nodes are the nodes that have no descendant nodes that are in the set  $P$ . This definition has a straightforward description in pseudocode. In the pseudocode in Fig. 7, the algorithm iterates over all problems in  $P$  and returns those problems whose sets of descendants, according to the ontology, have no elements in common with  $P$ .

After each iteration of the outer loop body of GETPROBLEMACTIVATIONSTRENGTHS, the levels of the activation strengths are updated by the UPDATEPROBLEMLEVELS algorithm. In the pseudocode of UPDATEPROBLEMLEVELS in Fig. 8, the algorithm first sets  $U$  to be the set of all advice units that are associated with active nodes in  $N$ . Then, for each advice unit, the algorithm tries to decrease the level of all problems that the advice unit applies to (i.e., all problems that are associated with the advice unit and all descendants of those problems). Some bookkeeping is done in  $A$  to ensure that one advice unit does not decrease the level of a node more than once (which could occur over the span of multiple iterations).

## 6.3. Calculating the advice unit priorities

In the second step of our *advice unit priority algorithm*, we convert activation strengths into advice unit priorities. The advice unit priorities map advice units to  $\langle \text{level}, \text{severity} \rangle$  tuples which, like the activation strengths, are ordered lexicographically by highest level first and by highest severity second. In fact, we define the *priority* of an advice unit as the maximum activation strength of the problems that are associated with the advice unit. The algorithm GETADVICEUNITPRIORITIES, shown in Fig. 9, shows a straightforward description of this definition and returns a mapping of advice units to priorities. These advice units are all the applicable advice units for the patient, based on the questionnaire answers provided, and the priorities are used to order the advice units.

From the algorithms used for our *advice unit priority algorithm*, we deduce that our approach ranks specific advice before generic advice and aims to diversify the top results (i.e., not letting the three advice units on the first page of advice all correspond to the same problem). For every advice unit associated with a problem in  $N$ , the UPDATEPROBLEMLEVELS algorithm decreases the level of the activation strengths of all problems that the advice unit applies to. Decreasing the levels of the activation strengths causes the affected problem nodes to have lower activation strengths for triggering advice in later iterations. We assume that the advice selected in later iterations is more generic, since it is associated with problem nodes that are more generic (because we traverse leaf nodes first, and leaf nodes are the most specific nodes according to the hierarchy of the ontology). Thus, by lowering the activation strengths of selected nodes after each iteration, our approach awards the highest rank to the most specific advice for a problem. Moreover, any advice triggered by the same problem in a later iteration is ranked lower than all specific advice (i.e., advice units triggered with an activation strength with level 0), regardless of severity.

Thus far, we assumed that there was one single filled-out questionnaire; however, our approach also works for multiple filled-out questionnaires. The only additional complication is that there now is a possibility that items of different questionnaires point to the same problem concept in the ontology. If this is the case, we take

**Algorithm 3:** UPDATEPROBLEMLEVELS( $N, T, A$ )

**Input:** set of problems  $N$ , associative array  $T$  mapping problems to  $\langle \text{level}, \text{severity} \rangle$  tuples, associative array  $A$  mapping problems to lists of advice units.  
**Data:** ontology function `descendants`, function `problems_associated_with`, function `advice_associated_with`.  
**Output:** updated  $T$  and  $A$ , where the mappings have been updated to reflect advice given by  $N$ .

$U \leftarrow$  empty set  
**for each** problem  $p \in N$   
  **do if**  $T[p]$   
  **then**  $\left\{ \begin{array}{l} \text{for each advice unit } a \in \text{advice\_associated\_with}(p) \\ \text{do add } a \text{ to } U \end{array} \right.$   
  **for each** advice unit  $u \in U$   
  **do**  $\left\{ \begin{array}{l} \text{for each problem } p \in \text{problems\_associated\_with}(u) \\ \text{do } \left\{ \begin{array}{l} \text{for each problem } q \in (\{p\} \cup \text{descendants}(p)) \\ \text{do if } T[q] \text{ and not } u \in A[q] \\ \text{then } \left\{ \begin{array}{l} \langle l, s \rangle \leftarrow T[q] \\ T[q] \leftarrow \langle l-1, s \rangle \\ A[q] \leftarrow A[q] \cup \{u\} \end{array} \right. \end{array} \right. \end{array} \right.$   
**return** ( $T, A$ )

Fig. 8. The UpdateProblemLevels algorithm.

**Algorithm 4:** GETADVICEUNITPRIORITIES( $B$ )

**Input:** associative array  $B$  mapping problems to  $\langle \text{level}, \text{severity} \rangle$  tuples (i.e., GETPROBLEMACTIVATIONSTRENGTHS()).

**Data:** function `advice_associated_with`.

**Output:** associative array mapping advice units to  $\langle \text{level}, \text{severity} \rangle$  tuples.

$R \leftarrow$  empty associative array

**for each** problem  $p \in B.\text{keys}$

**do** { **for each** advice unit  $a \in \text{advice\_associated\_with}(p)$

**do**  $R[a] \leftarrow \max(R[a], B[p])$

**return** ( $R$ )

**Fig. 9.** The GetAdviceUnitPriorities algorithm.

the (normalized) average of those answers as the problem severity for that problem.

#### 6.4. An example run

We will now illustrate the operation of the pseudocode of our *advice unit priority algorithm* by calculating advice priorities in an example scenario, shown in Fig. 10. This figure shows a subset of the nodes from Fig. 4, with the addition of an advice unit associated with the *School or work stress* node. In Fig. 10, as in Fig. 4, nodes with a black background are associated with advice units, nodes with a gray background are active nodes (i.e., associated with a questionnaire item that was answered above a certain threshold), and nodes with a white background are inactive and can be ignored. In this sample run, we will refer to the three nodes in Fig. 10 as  $\alpha$ ,  $\beta$ , and  $\gamma$ . Each of these nodes is associated with an item of the OQ-45 questionnaire, but only two nodes are considered active. We consider nodes as active only if they have a problem severity above a certain threshold (here we used 0.5). We will explain our motivation for using this particular threshold in more detail in the next section. For now, it is sufficient to know that we consider nodes  $\alpha$  and  $\gamma$  (with problem severities 0.67 and 0.75, respectively) as active and node  $\beta$  as inactive. Furthermore, note that node  $\alpha$  is the only node associated with an advice unit ( $\varphi$ : “Talk to case manager”).

The function GETPROBLEMACTIVATIONSTRENGTHS (from Fig. 6) is called with  $V = \{\alpha \Rightarrow 0.67, \gamma \Rightarrow 0.75\}$ . The node  $\beta$  is not included in  $V$  because it is not considered active. The variable  $P$  is initialized to  $P = \{\alpha, \beta, \gamma\}$  because it is simply a list of all nodes in the ontology. The variables  $B$ ,  $T$ , and  $A$  are initialized to empty associative arrays. The first for-loop sets  $T = \{\alpha \Rightarrow \langle 0, 0.67 \rangle, \gamma \Rightarrow \langle 0, 0.75 \rangle\}$ .

In the first iteration of the while-loop, we find as leaf nodes  $N = \{\beta, \gamma\}$ . Since neither of these nodes has descendants,  $T$  remains unchanged in the first inner loop.  $B$  becomes  $\{\gamma \Rightarrow \langle 0, 0.75 \rangle\}$ . Note that  $\beta$  is not included in  $B$  because  $\beta$  was not included in  $V$ . Variables  $T$  and  $A$  remain unchanged after the call to UPDATEPROBLEMLEVELS

(from Fig. 8), since none of the nodes in  $N$  are associated with advice units.

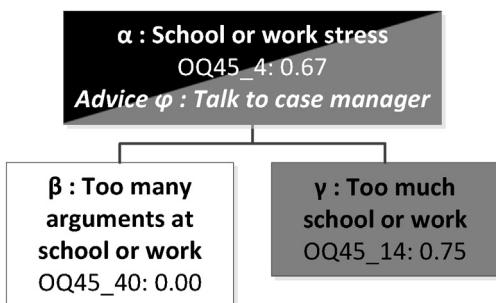
In the second iteration of the while-loop in GETPROBLEMACTIVATIONSTRENGTHS, by having removed  $\beta$  and  $\gamma$  from  $P$ , we now find  $N = \{\alpha\}$ , and  $T$  becomes  $\{\alpha \Rightarrow \langle 0, 0.75 \rangle, \gamma \Rightarrow \langle 0, 0.75 \rangle\}$ , since  $\gamma$  is a descendant of  $\alpha$ . These are also the values returned by  $B$ . After the second iteration, UPDATEPROBLEMLEVELS sets  $A$  to  $\{\alpha \Rightarrow \varphi, \gamma \Rightarrow \varphi\}$  and  $T$  to  $\{\alpha \Rightarrow \langle -1, 0.75 \rangle, \gamma \Rightarrow \langle -1, 0.75 \rangle\}$ , signifying that an advice unit  $\varphi$  was given that applies to these problems. These values for  $T$  would normally be used in future iterations; however, in this example, there are no future iterations, since there are no nodes left in  $P$ .

The second step in our approach in Fig. 5 is to call the function GETADVICEUNITPRIORITIES (from Fig. 9) with  $B = \{\alpha \Rightarrow \langle 0, 0.75 \rangle, \gamma \Rightarrow \langle 0, 0.75 \rangle\}$ . Since the only node associated with an advice unit in our example is node  $\alpha$ , and since this node is included in  $B$ , we find that this results in  $R = \{\varphi \Rightarrow \langle 0, 0.75 \rangle\}$ .

Thus, for this sample scenario we find that the list of selected advice units consists of a single advice unit  $\varphi$  triggered with priority  $\langle 0, 0.75 \rangle$ . The level 0 signifies that the advice unit is the most specific advice unit for a certain problem (School or work stress, i.e., node  $\alpha$ , for which the strength is calculated as the maximum of it and its descendants that are not covered by a more specific advice unit) and that it should be sorted by severity among other level 0 advice units, that is, before any advice units triggered with level  $-1$  or lower. In the next section, we will validate and test our approach against the opinions of clinicians and patients.

## 7. Experiments and results

We will evaluate the utility of our system in two experiments, both based on results of the MANSA questionnaire [3]. The first experiment compares the identification of important problems vis-à-vis the opinions of clinicians, and the second experiment compares the selection of relevant advice topics vis-à-vis the opinions of patients. For our first experiment, given a set of filled-out questionnaires, we tested how closely our method which is based on problem severities corresponds, in terms of identifying important problems, to the opinions of clinicians who give patients advice on a day-to-day basis. The goal is to determine whether clinicians are primarily steered by the type of problem (i.e., some problems are considered more important than others) or by the severity of the problem, our system being based on the latter assumption. For our second experiment, we measure the effects of using a *severity threshold* to truncate the list of advice units for a patient by letting patients evaluate the perceived relevance of selected advice topics. Additionally, this experiment allows us to draw conclusions about whether the system is considered helpful and relevant by the patients. We chose to use the MANSA questionnaire for our experiments because: (i)

**Fig. 10.** An example scenario with three nodes.

it is part of the standard ROM protocol; (ii) it is a relatively short questionnaire, yet it identifies a variety of problems; and (iii) it can be filled out by the patients themselves. In the following section, we will introduce some concepts common to both experiments.

### 7.1. Evaluation measurements

In the evaluation of the results of our experiments, we used measurements of precision, recall, and their harmonic mean (also called the *F-measure*). In both experiments, for each filled-out questionnaire, we compared two selections, one made by the system and one made by the expert. We established the selection made by the expert as a *ground truth*, allowing the relevance of the selection made by the system to be expressed in terms of precision, recall, and harmonic mean. The *precision* is the fraction of items selected by the system that are also selected by the expert, while *recall* is the fraction of items selected by the expert that are also selected by the system. We applied these measurements in both experiments, but we applied them to different concepts. The selections made by the system and experts consist of items (called “topics” in the formulas below), which are problem areas for our first experiment and advice units for our second experiment. Likewise, the term “expert” refers to the clinicians for our first experiment and to the patient for our second experiment. Furthermore, the *selections* are the topics considered most relevant.

We calculated the precision, recall, and harmonic mean using a cut-off to consider only the first  $n$  topics ( $n = 1, 2, 3$ ). The first three topics form a good evaluation criterion for our experiments, since Wegweis shows only three advice units on the first page of advice for a patient. In the following definitions, let  $T_n^e$  denote the set of the  $n$  most relevant topics according to the expert, and let  $T_n^s$  denote the set of the  $n$  most relevant topics according to the system. We formulate  $P_n$  (i.e., precision at  $n$ ) as follows [40].

$$P_n = \frac{t \in \{T_\infty^e \cap T_n^s\}}{t \in T_n^s}$$

Here,  $t$  denotes the number of topics. Thus, precision at  $n$  is the fraction of the  $n$  most relevant topics identified by the system that are also identified as relevant by the expert. Likewise, we define  $R_n$  (i.e., recall at  $n$ ) as follows [40].

$$R_n = \frac{t \in \{T_n^e \cap T_\infty^s\}}{t \in T_n^e}$$

Thus, recall is the fraction of the  $n$  most relevant topics identified by the expert that are also identified as relevant by the system. Finally, we define  $F_n$  (i.e., the harmonic mean of precision and recall at  $n$ ) as follows.

$$F_n = 2 \cdot \frac{P_n \cdot R_n}{P_n + R_n}.$$

In our experiments, we evaluated the effects of applying a *severity threshold* to limit the number of results returned. If we were to simply return all results, that is, marking as relevant every problem that did not have a perfect answer, the patient would be overwhelmed by the amount of advice and would receive a lot of advice for issues that he/she would not consider to be a problem (e.g., MANSA items answered with 6 = “Pleased”). Thus, since we base our relevance selection solely on problem severity, we needed to use a severity threshold to limit the amount of results returned. The MANSA questionnaire consists of 16 items, 4 of which are binary items (i.e., answered using “Yes” or “No”) and the other 12 are rated on a seven-point satisfaction scale (ranging from 1 = “Couldn’t be worse” to 7 = “Couldn’t be better”). Since the most complex answer type in the MANSA questionnaire is a seven-point rating scale, there are six possible thresholds. To find the best threshold, we evaluated

these described measurements for all threshold values on our test set. The results listed “with thresholding” correspond to the optimal threshold value (which ignores answers in the 5–7 range).

In cases where there is no unique ordering (e.g., because multiple problems have the same severity), we take the average over all possible permutations that satisfy the criterion of being sorted according to severity. This guarantees that the ordering depends solely on severities, even when these are equal, without introducing an arbitrary bias.

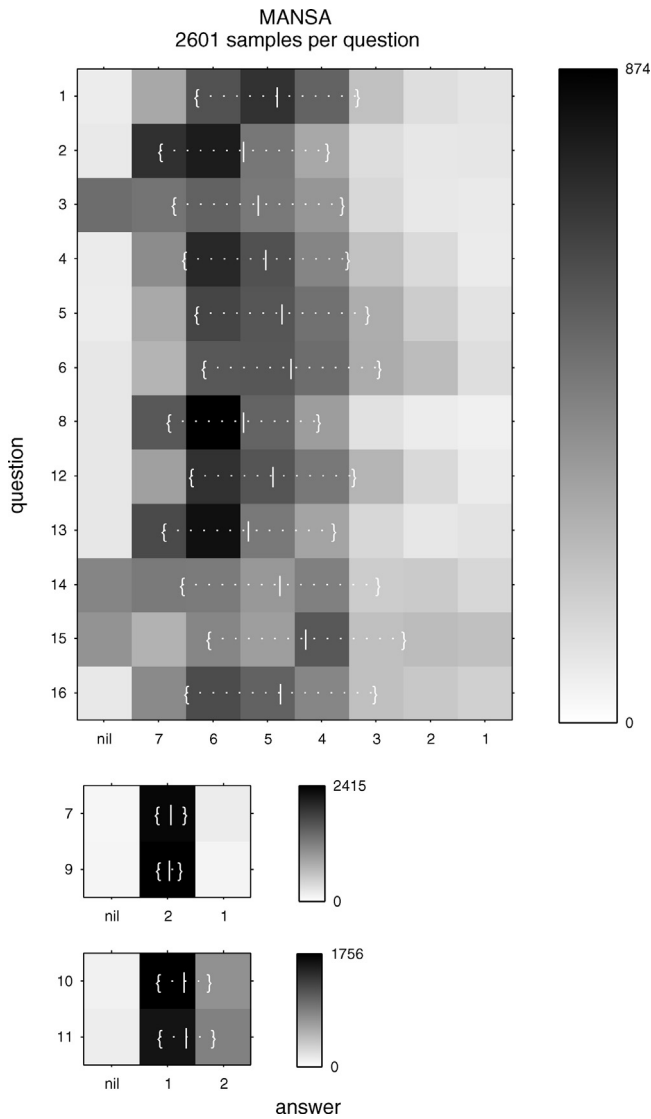
### 7.2. Clinicians and problem severities

As our first experiment, we will test how a system based on problem severities corresponds to the opinion of clinicians, with respect to identifying important problems in the MANSA questionnaire. We executed this experiment twice, with different sets of samples, and the results presented in this section pertain to the two sets combined. In the first execution, we selected five samples (i.e., filled-out MANSA questionnaires) with several severe problems and asked five clinicians (2 psychiatrists and 3 nurse practitioners) to give a list of problem areas in descending order of importance, which they would discuss with the patient, for each sample. We then compared these 25 results to those of Wegweis. In the second execution, we repeated this experiment with 3 clinicians and 30 samples. Contrary to the first set of samples, this second set was chosen fully at random, that is, the samples did not necessarily have any severe problems. In point of fact, five of the samples in this set actually did not have any severe problems. The executions amounted to a total of 35 samples, which were evaluated by clinicians in 115 lists, which we then compared with the results of Wegweis. The samples that we used in this experiment were selected from a data set (which we acquired through Roqua) of MANSA questionnaires filled out by schizophrenia patients.

Five of the samples that we used in the second execution for this experiment did not include any severe problems and so were excluded from this test. The reason for this was that we cannot use samples without severe problems to prove or disprove our assumption that clinicians select severe problems. Moreover, with severity thresholding applied, our approach only gives results for a sample when it contains severe problems. From our data set of 2601 samples from 1379 patients, 291 samples (11.19%) had no severe problems. We simply accepted the fact that our approach did not apply to the 11.19% of schizophrenia patients who had no severe problems, which we justify by arguing that we do not need to give advice if there is no need for it.

An impression of the distribution of answers of schizophrenia patients for this questionnaire is given in Fig. 11. This figure shows 2601 filled-out MANSA questionnaires from 1379 schizophrenia patients in the Northern Netherlands as *heat maps*. A heat map is a two-dimensional plot in which the values of a variable are embedded through color intensities or gray levels. In Fig. 11, the gray level denotes the sample frequency, such that the average gray level of each row is the same, that is, dark squares denote popular choices. The figure shows three heat maps, one for each answer type of the MANSA. The severity of the responses increases from left to right, with the two smaller heat maps representing the yes/no and no/yes items. The braces give an indication of the spread of the answers for an item, and are placed at one standard deviation from the mean on either side. The *nil* column indicates missing or blank values, which are ignored. This figure shows that even though the questionnaire has only 16 questions, many distinct combinations of answers exist, and identifying the important problems is not a trivial task.

We established the *ground truth* in this experiment by averaging over the rankings given by the clinicians. For each sample, this resulted in a single ordered list of problem areas. However, these lists could include outliers (e.g., topics that were selected by



**Fig. 11.** Heat map showing answers from schizophrenia patients in 2601 MANSA questionnaires.

only one clinician) that should be discarded. For this purpose, we restricted the maximum length of the list of topics selected by the clinicians to the number of severe problems in the sample. Our reason for basing the cut-off on the number of severe problems is that we are interested in the problems that are considered relevant by clinicians in spite of other problems that are more severe. For example, if a sample indicates three severe problems, and we consider the first three problems selected by the clinicians as relevant, then any difference with the selection of the system is an indication of non-severe problems that clinicians consider more relevant than certain severe problems.

We compared the selections of the clinicians to the selections of the system with thresholding, and the result is shown in Table 1. This table shows measurements of precision, recall, and *F*-measure

**Table 1**  
Comparing the system (with thresholding) to the opinion of the clinicians.

<i>n</i>	Precision @ <i>n</i>	Recall @ <i>n</i>	<i>F</i> -measure @ <i>n</i>
1	0.983	1.000	0.992
2	0.957	1.000	0.978
3	0.943	0.944	0.944

**Table 2**

A breakdown per topic for  $n=\infty$ , comparing the system (with thresholding) to the opinion of the clinicians.

Topic	Only clinicians	Only system	Both
Sex	0.0% (0)	66.7% (12)	33.3% (6)
Physical health	0.0% (0)	38.5% (5)	61.5% (8)
Daily activities	30.8% (4)	7.7% (1)	61.5% (8)
Life	8.3% (1)	25.0% (3)	66.7% (8)
Security	18.8% (3)	12.5% (2)	68.8% (11)
Finances	0.0% (0)	28.6% (4)	71.4% (10)
Housing	5.3% (1)	10.5% (2)	84.2% (16)
Psychic health	11.8% (2)	0.0% (0)	88.2% (15)
Relationships	0.0% (0)	7.7% (2)	92.3% (24)
Accused of crime	0.0% (0)	0.0% (0)	100.0% (2)

for  $n = 1, 2, 3$ . From Table 1 we note that with severity thresholding we retain perfect recall values for  $n = 1$  and  $n = 2$ . Thus, we find that in our experiments, the two most important topics according to a clinician are always severe problems. Moreover, for the first three results, our approach based on problem severities complies with clinicians evaluations on average 94% of the time.

While Table 1 shows the similarity between system and clinicians for the first three results, for a comparison of the full selections (i.e., for  $n=\infty$ ), we refer to Table 2. This table gives a breakdown per topic of the selections made by system and clinicians. The “Only clinicians” column shows the topics that were non-severe problems yet were included by clinicians, the “Only system” column shows the problems that were severe yet were excluded by clinicians, and the “Both” column shows topics that were included by both. On average, we find that 7.3% of selected topics were non-severe problems yet were included by clinicians, and 20.7% were severe problems yet were excluded by clinicians. Thus, for the full selections, our approach corresponds 72.0% of the time with the clinicians, but as we saw in Table 1, this percentage is higher (94%) for the first three results.

### 7.3. Patients and advice relevance

For our second experiment, we evaluated to what extent the advice units selected by Wegweis for a patient were considered relevant by that patient. In this experiment, we let patients fill out a MANSA questionnaire and had them evaluate the advice selected by the system, based on those questionnaire answers. We performed this particular experiment for two reasons. First, this experiment allows us to evaluate the effect, with respect to patient satisfaction, of limiting the number of selected advice units by applying a severity threshold. We evaluated this effect by presenting the patients with all the applicable advice units, letting them make their own selection of relevant advice, and then comparing that selection to the selection of the system after applying the severity threshold. Second, this experiment evaluated our advice selection and the ranking algorithms that were explained in Section 6. These algorithms are used because the connection between questionnaire items and advice units is not necessarily direct but can be inferred through the problem ontology. Thus, the advice selection for a patient can, for instance, contain very generic advice for very specific problems. Therefore, the assumption to be tested is that the overall selection of advice is still deemed relevant by the patient.

In this experiment, the *ground truth* is the opinion of the patient who filled out the questionnaire, and the results are averaged over all patients. For this experiment, we asked 13 patients (for information on the selection procedure for patients, we refer to our usability study [29]) to fill out the MANSA questionnaire. These filled-out questionnaires were then processed by Wegweis to calculate the full set of applicable advice units (i.e., without thresholding) for each patient. The patients were then asked to select from their set



**Table 3**

Comparing the system (with and without thresholding) to the opinion of the patients.

<i>n</i>	Precision @ <i>n</i>	Recall @ <i>n</i>	<i>F</i> -Measure @ <i>n</i>
Without thresholding			
1	0.652	1.000	0.790
2	0.617	1.000	0.763
3	0.665	1.000	0.798
∞	0.361	1.000	0.530
With thresholding			
1	0.652	0.846	0.737
2	0.643	0.808	0.716
3	0.702	0.815	0.754
∞	0.574	0.756	0.653

those advice units that they considered relevant to their personal situation and to list them in order of relevance. We told the patients to evaluate the relevance of the topics of the advice units (i.e., the advice titles) and not the relevance of the advice contents. The advice contents were not evaluated in this paper, because they were independent of our approach for inferring, selecting, and ranking advice.

The results of comparing the selections of the patients to the selections of the system (both with and without thresholding) are shown in Table 3. This table shows measurements of precision, recall, and *F*-measure for  $n = 1, 2, 3, \infty$ . The thresholding used for the bottom half of the table is the same thresholding we used in our first experiment, that is, it implies that the system will ignore non-severe problems. The perfect (1.000) values for recall in the top half of Table 3 are explained by the fact that the system does not omit any advice unless a threshold is used.

In Table 3, we find that for increasing values of  $n$ , the measurements do not show a steady decrease but show fluctuation. This fluctuation is due to the fact that the measurements for different values of  $n$  are based on different amounts of samples, because some samples have only one or two relevant advice units. For example, when the number of relevant advice units for a sample according to the system (or the patient) is two, then this sample will be included in the average for  $n = 2$  but not in the average for  $n = 3$ . Despite these fluctuations, we can derive that, for our advice system based on severities, on average two of the three advice units on the first page of advice are considered relevant by the patient (0.702 precision at  $n = 3$ ).

Table 3 also shows that applying a severity threshold results in a higher *F*-measure when comparing all relevant advice. The rows with  $n = \infty$  in Table 3 correspond to the standard definitions for precision, recall, and *F*-measure. These rows show that the precision increases when applying a severity threshold. More specifically, when applying a threshold, 57.4% of the advice given is considered relevant by patients, up from 36.1%. This increase in precision comes coupled with a decrease in recall from 100% to 75.6%, which indicates that only 75.6% of the advice units considered relevant by the patients link to severe problems. However, the combined effect of thresholding remains positive. This effect is shown by the increase of *F*-measure (from 0.530 to 0.653). These findings suggest that, according to the patients, the use of the severity threshold improves the quality of the advice returned by the system. A breakdown into individual advice topics was omitted from this paper, since it did not identify any significant trends.

The values of Table 3 are relatively low, which indicates that, for patients, the problem severity is not the only criterion for determining the relevance of an advice unit. For example, in our experiment, there were multiple patients with severe problems who marked only non-severe advice units as relevant. In a dismissed alternative approach, we applied global relevance learning to identify popular advice units for patients. However, we found that global relevancies

did not improve the results. This outcome suggests that the relevant advice selection of patients is highly patient-specific.

We performed a second run of the experiment by inviting another 14 patients to use and evaluate our system, to comment on its utility, and to report any abnormalities. Their responses were consistent with our earlier observations. Eight patients responded to our invitation, five of whom had severe problems. For these five patients, of the first three advice units selected by the system with thresholding, 46.7% was found relevant. A possible explanation as to why this number is lower is because, for this run, we used questionnaire data from the most recent assessment of the patients, which was outdated in some cases. For example, one patient remarked that the advice addressed problems that he had reported six months earlier but which had been resolved since then, and thus the associated advice was no longer relevant. In a typical setting, where Wegweis is used as soon as the assessment results are in, the relevance is likely to be higher.

## 8. Discussion

Prior studies have noted the importance of ethical imperatives such as shared decision making [7]. Shared decision making requires the sharing of medical information between patient and clinician. In the current treatment of schizophrenia patients, the clinician decides which information is shared. We believe that information sharing and shared decision making as a whole can be facilitated by automated ways of interpreting and explaining medical data in forms that are accessible and understandable for patients.

The results of our current study show that for the task of identifying the most important problems from a filled-out MANSA questionnaire, an approach based on problem severities can be an adequate approximation of the way clinicians prioritize information for a patient. For the three most important problems, our approach corresponded to the opinion of clinicians in 94% of tested cases, and for all problems, our approach corresponded in 72%. The differences appear to be restricted to a subset of the topics. For example, in Table 2, we find that frequently occurring problems such as housing, psychic health, and relationships were identified by the system and clinicians roughly equally often. However, sexual problems, finances, and physical health are issues that clinicians sometimes choose to omit, even when these problems are severe. In contrast, clinicians sometimes discuss daily activities without these being a severe problem. The possible bias for this topic was explained by one of the clinicians, who remarked that when there is nothing else to discuss, they would ask the patient what their plans were for the upcoming week, which is a discussion topic that would be classified under daily activities in our experiments. Another clinician remarked that they would ask the patient if they had any other problems or topics that they wanted to discuss. While not modeled in the results, this interaction roughly equates to the search function on the Wegweis website.

However, we found that patients do not prioritize information in the same way as clinicians do (i.e., using only problem severities). While problem severities have some significance for patients, patients, in their relevance selections, may consider other factors which are unknown to us. In spite of this fact, our experiments show that patients still consider most advice given by the system to be relevant and perceive a quality improvement when a severity threshold is used. The fact that the severity threshold had a positive effect was explained during our feedback sessions by patients, who stated that they did not appreciate being given advice for problems where they had answered 6 = "Pleased" instead of 7 = "Couldn't be better." Our experiments also tested the use of the problem ontology to infer generic advice for specific problems, since 5 of the



16 MANSA items had no directly associated advice in the problem ontology at the time of testing. Inferring advice through the ontology did not lead to any logically unexpected advice, according to the patients. Feedback from patients concerning the relevance of advice was related mostly to the contents of the advice rather than to the reason that the advice was given. For example, one patient noted that he talked about physical problems with his physician and not his psychiatrist.

The results of this study are consistent with those of other studies that demonstrated the utility of self-management applications in healthcare [15]. Furthermore, our experiments have not yielded any evidence to support the traditional belief that there is danger in giving schizophrenia patients direct access to their medical information. On the contrary, our experiments are consistent with the more recent belief that patients benefit from shared decision making [5].

The results need to be interpreted with caution as they are based on small sample sizes. Moreover, our approach only applies for samples that have at least one severe problem, otherwise no advice is shown. Furthermore, the experiment with clinicians is not an entirely accurate scenario in some cases, since in practice clinicians will take the patient history into account when giving advice. Whether or not this would shift the results significantly and whether the patient would benefit more from biased or unbiased advice are topics of debate.

Our findings suggest that an approach based on problem severities is adequate for identifying important problem areas from schizophrenia-related questionnaires, and that such an approach can be considered helpful and relevant by patients in selecting and ranking advice.

These findings have important implications for the development of systems that automate the translation and interpretation of assessment results for patients with chronic illnesses. If such systems can be shown to work for schizophrenia patients, who impose numerous restrictions on the user interface, then these systems are likely to work for patients with other chronic illnesses too. In those branches of healthcare, this paves the way for automated solutions that support the sharing of information between patient and clinician as an integral part of shared decision making.

The present results are significant because they demonstrate the efficacy of an intuitive way to prioritize information in the same way as a clinician would. However, our approach does not explain the relevance selection of the patients very well, leaving room for improvement.

## 9. Conclusions and future work

We have presented the development, the design, the testing, and the evaluation of Wegweis, a patient-centered web application driven by an ontology-based approach that uses ROM assessment results to select and rank advice for schizophrenia patients. The system has minimal impact on the way clinicians work, because it integrates with an existing questionnaire manager. Adding support for a questionnaire in Wegweis is simplified by the fact that questionnaires are decoupled from advice by virtue of the problem ontology. Background knowledge, embedded in the structure of the ontology, is used to infer advice when no exact match is found, which adds to the robustness of the system.

The study set out to determine whether a fully automated explanation and interpretation of ROM assessment results for schizophrenia patients that prioritizes the information in the same way that a clinician would is possible, and whether it would be considered helpful and relevant by patients. The evidence from this study suggests that such an automated explanation and interpretation is indeed possible and considered relevant by

patients, and thus can be a helpful addition in improving patient care. The improvement is due to two reasons. First, an automated explanation and interpretation of assessment results empowers the patient because it allows patients to prepare for discussing their treatment plan without requiring any help. Second, where clinicians may forget to mention or choose to ignore certain alternatives, an automated approach presents the patient with all the options it knows about and leaves the decision up to the patient. We conclude that a system such as Wegweis can work as a useful adjunct to the care of schizophrenia patients in the form of a *second perspective*: unbiased advice that is ordered in a way that has high similarity to what a clinician would discuss, given the same questionnaire data.

The approach we used for selecting and ranking advice can be used to enhance self-management websites for other chronic illnesses as well. Since all domain knowledge is stored in the ontology, the approach lends itself to providing personalized advice in other areas of healthcare.

Finally, a number of important limitations need to be considered. First, an advice system relies heavily on the domain-specific problem ontology and on the advice contents. Moreover, its performance is very dependent on the specific questionnaires. Thus, porting the approach to other areas of healthcare would not be a trivial task. A new ontology would have to be built, based on disease-specific questionnaires and terms, and a new body of advice contents would have to be collected and validated by experts. Second, the main weakness of our study was the small number of patients who evaluated the advice selections of our system, and those results may therefore not be transferable to schizophrenia patients in general.

Our research has raised many questions in need of further investigation. More experiments are needed to determine how questionnaires other than the MANSA would score in the experiments. Another issue worth investigating is the extent to which clinicians take the patient history into account when identifying important problems, and how this can be modeled. Another unaddressed question is how to make the advice rankings match the patient opinions more closely. An approach that takes previous assessments into account may help to construct a more complete image of a patient and would allow for reasoning over changes in the condition of a patient over time. While we are aware that some work has been started in this area [22], we believe that these efforts could benefit from an ontology-based approach.

## Acknowledgements

The research is supported by the Dutch National Research Council for Innovation in Health Care (ZonMW), under contract no. 300020011. We thank the anonymous reviewers for their constructive criticism.

## References

- [1] Van Os J, Sham P. Gene–environment interactions. In: Murray R, Jones P, Susser E, van Os J, Cannon M, editors. The epidemiology of schizophrenia. Cambridge, UK: Cambridge University Press; 2003. p. 235–54.
- [2] Wing J, Beevor A, Curtis R, Park S, Hadden S, Burns A. Health of the Nation Outcome Scales (HoNOS): research and development. The British Journal of Psychiatry 1998;172(1):11–8.
- [3] Priebe S, Huxley P, Knight S, Evans S. Application and results of the Manchester Short Assessment of Quality of Life (MANSA). International Journal of Social Psychiatry 1999;45(1):7.
- [4] Deegan P, Drake R. Shared decision making and medication management in the recovery process. Psychiatric Services 2006;57(11):1636–9.
- [5] Godolphin W. Shared decision-making. Healthcare Quarterly 2009;12:e186–90.
- [6] Woltmann E, Wilkiss S, Teachout A, McHugo G, Drake R. Trial of an electronic decision support system to facilitate shared decision making in community mental health. Psychiatric Services 2011;62(1):54–60.
- [7] Drake R, Deegan P. Shared decision making is an ethical imperative. Psychiatric Services 2009;60(8):1007.

- [8] Charles C, Gafni A, Whelan T. Shared decision-making in the medical encounter: what does it mean? (or it takes at least two to tango). *Social Science & Medicine* 1997;44(5):681–92.
- [9] Schrank B, Sibitz I, Unger A, Amering M. How patients with schizophrenia use the internet: qualitative study. *Journal of Medical Internet Research* 2010;12(5):e70.
- [10] Paganelli F. An ontology-based context model for home health monitoring and alerting in chronic patient care networks. In: O'Conner L, editor. *Proceedings of the 21st international conference on advanced information networking and applications workshops*, vol. 2. IEEE Press; 2007. p. 838–45.
- [11] Buranarach M, Supnithi T, Chalortham N, Khunthong V, Varasai P, Kawtrakul A. A semantic web framework to support knowledge management in chronic disease healthcare. In: Sartori F, Sicilia M, Manouselis N, editors. *Metadata and semantic research*; vol. 46 of communications in computer and information science. Berlin, Heidelberg: Springer; 2009. p. 164–70.
- [12] Patel C, Gomadam K, Khan S, Garg V. TrialX: using semantic technologies to match patients to relevant clinical trials based on their Personal Health Records. *Web Semantics: Science, Services and Agents on the World Wide Web* 2010;8(4):342–7.
- [13] Maier E, Reimer U, Schär S, Zimmermann P. SEMPER: a web-based support system for patient self-management. In: Owens T, editor. *Proceedings of the 23rd bled eConference*, vol. 17. Atlanta: AIS Electronic Library; 2010. p. 196–209.
- [14] Kuriyama D, Izumi S, Itabashi G, Kimura S, Ebihara Y, Takahashi K, et al. Design and implementation of a health management support system using ontology. In: Chotikakamton N, editor. *Proceedings of the international conference on engineering, applied sciences, and technology*. Thailand: IEEE; 2007. p. 746–9.
- [15] Proudfoot J. Computer-based treatment for anxiety and depression: is it feasible? Is it effective? *Neuroscience & Biobehavioral Reviews* 2004;28(3): 353–63.
- [16] Kersting A, Schlicht S, Kroker K. Internet therapy: opportunities and boundaries. *Der Nervenarzt* 2009;80(7):797.
- [17] Riper H. E-mental health: high tech, high touch, high trust. Utrecht: Trimbo Instituut; 2007.
- [18] Välimäki M, Anttila M, Hätönen H, Koivunen M, Jakobsson T, Pitkänen A, et al. Design and development process of patient-centered computer-based support system for patients with schizophrenia spectrum psychosis. *Informatics for Health and Social Care* 2008;33(2):113–23.
- [19] Koivunen M, Välimäki M, Pitkänen A, Kuosmanen L. A preliminary usability evaluation of web-based portal application for patients with schizophrenia. *Journal of Psychiatric and Mental Health Nursing* 2007;14(5):462–9.
- [20] Koivunen M, Välimäki M, Patel A, Knapp M, Hätönen H, Kuosmanen L, et al. Effects of the implementation of the web-based patient support system on staff's attitudes towards computers and its use: a randomised controlled trial. *Scandinavian Journal of Caring Sciences* 2010;24(3):592–9.
- [21] Kuosmanen L, Jakobsson T, Hyttinen J, Koivunen M, Välimäki M. Usability evaluation of a web-based patient information system for individuals with severe mental health problems. *Journal of Advanced Nursing* 2010;66(12):2701–10.
- [22] Eigen regie bij schizofrenie. <http://www.eigen-regie.nl>; 2011 [accessed 18 November 2012].
- [23] Deegan P, Rapp C, Holter M, Riefer M. Best practices: a program to support shared decision making in an outpatient psychiatric medication clinic. *Psychiatric Services* 2008;59(6):603–5.
- [24] Samen keuzes maken. <http://www.samenkeuzesmaken.nl>; 2011 [accessed 18 November 2012].
- [25] Roqua. <http://www.roqua.nl>; 2011 [accessed 18 November 2012].
- [26] Dolin R, Alschuler L, Boyer S, Beebe C, Behlen F, Biron P, et al. H17 clinical document architecture, release 2. *Journal of the American Medical Informatics Association* 2006;13(1):30–9.
- [27] Crockford D. The application/json media type for javascript object notation (json). Available from: <https://tools.ietf.org/html/rfc4627>; 2006 [accessed 18 November 2012].
- [28] Rotondi A, Sinkule J, Haas G, Spring M, Litschge C, Newhill C, et al. Designing websites for persons with cognitive deficits: design and usability of a psychoeducational intervention for persons with severe mental illness. *Psychological Services* 2007;4(3):202–24.
- [29] Van der Krieke L, Emerencia AC, Aiello M, Sytema S. Usability evaluation of a web-based support system for people with a schizophrenia diagnosis. *Journal of Medical Internet Research* 2012;14(1):e24.
- [30] Mobasher B, Cooley R, Srivastava J. Automatic personalization based on web usage mining. *Communications of the ACM* 2000;43(8):142–51.
- [31] Browsercms. <http://www.browsercms.org>; 2011 [accessed 18 November 2012].
- [32] Liquid templating language. <http://liquidmarkup.org>; 2011 [accessed 18 November 2012].
- [33] Trauer T, Tobias G, Slade M. Development and evaluation of a patient-rated version of the camberwell assessment of need short appraisal schedule (cansasp). *Community Mental Health Journal* 2008;44:113–24.
- [34] Lambert M, Finch A. The outcome questionnaire. In: Maruish ME, editor. *The use of psychological testing for treatment planning and outcomes assessment*. 2nd ed. Mahwah, NJ, USA: Lawrence Erlbaum Associates Publishers; 1999. p. 831–69.
- [35] Stearns M, Price C, Spackman K, Wang A. Snomed clinical terms: overview of the development process and project status. In: Bakken S, editor. *Proceedings of the American Medical Informatics Association Symposium*. Philadelphia, PA: Hanley & Belfus Inc.; 2001. p. 662–6.
- [36] Gennari JH, Musen MA, Fergerson RW, Grosso WE, Crubézy M, Eriksson H, et al. The evolution of protégé: an environment for knowledge-based systems development. *International Journal of Human-Computer Studies* 2003;58(1):89–123.
- [37] McGuinness D, Van Harmelen F. Owl web ontology language overview (w3c recommendation). <http://www.w3.org/TR/owl-features/>; 2004 [accessed 18 November 2012].
- [38] Wegweis ontology. Available from: <http://www.wegweis.nl/ontologies/problems.owl>; 2011 [accessed 18 November 2012].
- [39] Emerencia A, van der Krieke L, Petkov N, Aiello M. Assessing schizophrenia with an interoperable architecture. In: Bouamrane MM, Tao C, editors. *Proceedings of the first international workshop on managing interoperability and complexity in health systems*. New York, NY: ACM; 2011. p. 79–82.
- [40] Van Rijsbergen C. *Information retrieval*. 2nd ed. London: Butterworth & Co. Ltd.; 1979. p. 111–43 [chap. 7].